

¹Departamento de Producción Animal, Facultad de Veterinaria, Universidad Complutense, Madrid

²Departamento de Producción Animal, Escuela Técnica Superior de Ingenieros Agrónomos,
Madrid, España

Comparison of two models for estimation of variance components in a sample of Spanish Holstein Friesians

By J. P. GUTIÉRREZ,¹ J. CAÑON,¹ and M. RICO²

Introduction

Most research in animal breeding methodology has been undertaken in developed countries. In these, methodology has been applied to large data sets, more or less balanced and connected, and containing full genealogy. However, the situation in many other countries is really far from such type of data sets. Thus, the efficiency of some methods should be tested, before introducing more expensive techniques which could be unnecessary.

HENDERSON (1975) suggested three possible ways for comparing alternative methods. The first one is based on analytic methods, which require complicated mathematic manipulations. The second possibility is to check methods on simulated data, where true values are known (LIN and McALLISTER 1984; VAN DER WERF and DE BOER 1989). In this case, conclusions are dependent on the simulated population. Finally, the most extended way, is to apply them to the same real data set (DEMPFLE et al. 1983; VAN RADEN and JUNG 1988). If results are different it is not possible to say which method is better.

The objective of this research was to develop a simulation method which allows comparison of variance component estimation methods using different models in a particular population.

Materials and methods

Data

A sample of 5,496 Spanish Holstein Friesian lactations, covering the period 1980 to 1988 were used. Records belonged to 3,725 cows, daughters of 324 sires and distributed in 354 herds. The pedigree file contained 4,946 animals.

Unbalancedness, shown in Figures 1, 2 and 3, and Table 1, can be summarized as follows:

- More than one half of the herds were represented with 5 or less cows. Less than 5% had more than 35 (Figure 1).

- 66.5% of sires had between 5 and 10 daughters and only 1% of sires had more than 50 (Figure 2).

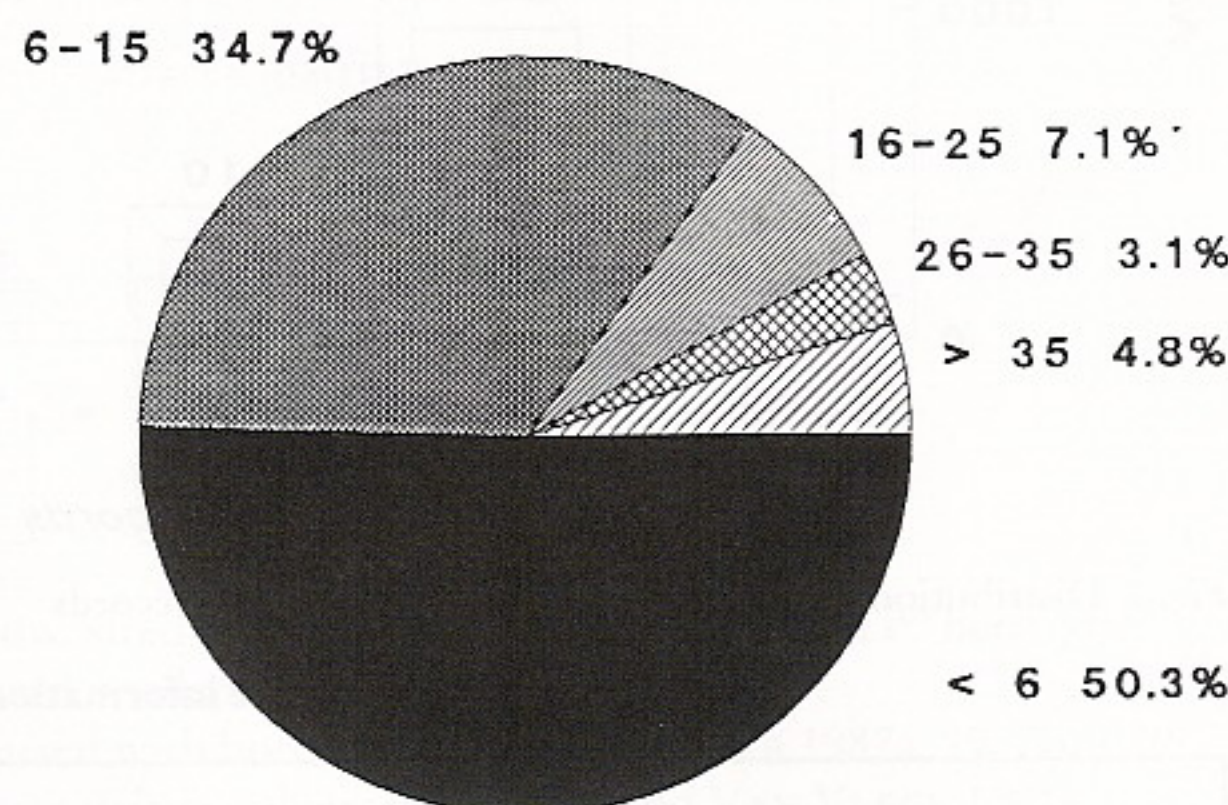


Fig. 1. Distribution of herds according to the number of cows

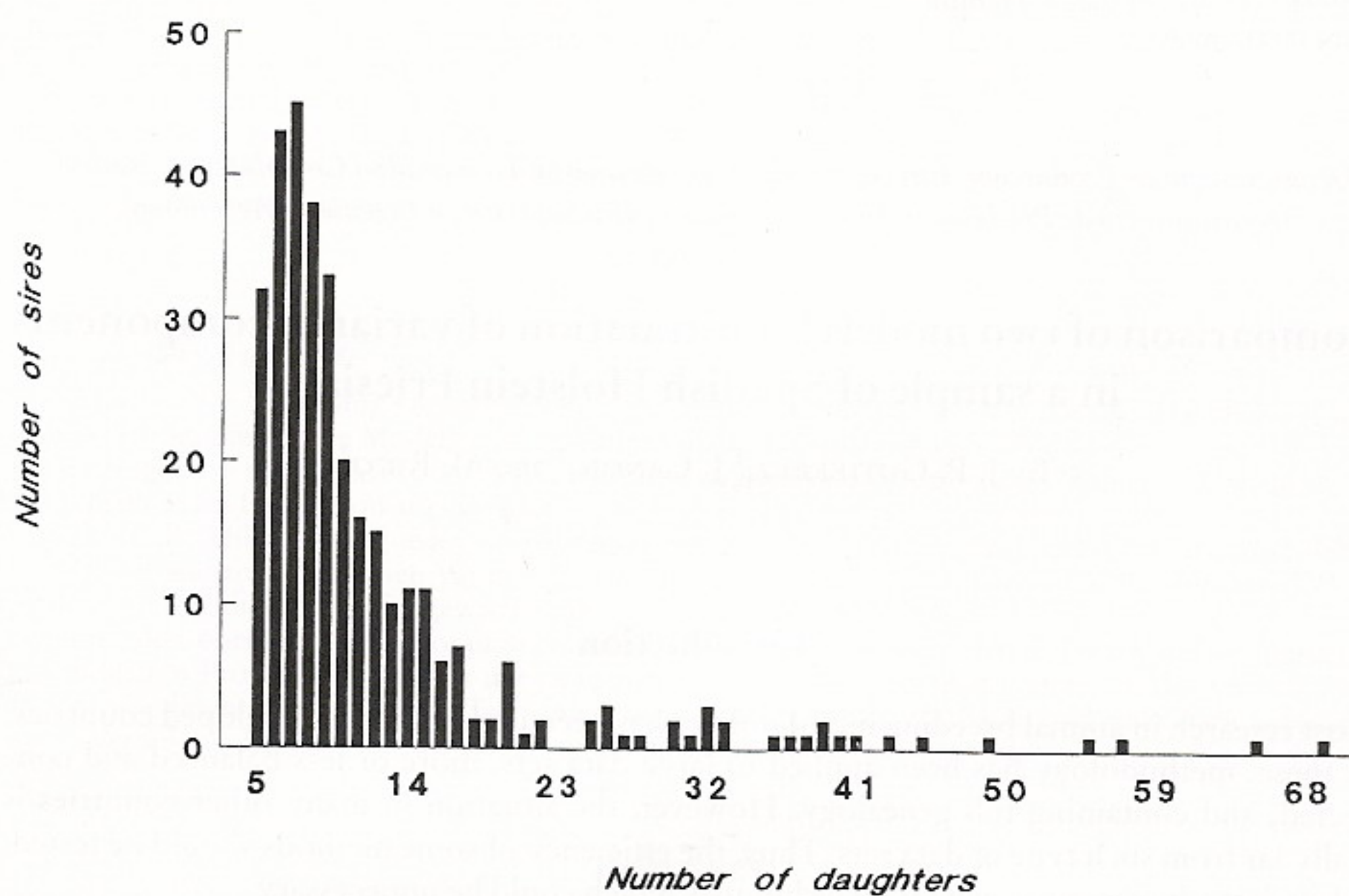


Fig. 2. Distribution of sires according to the number of daughters

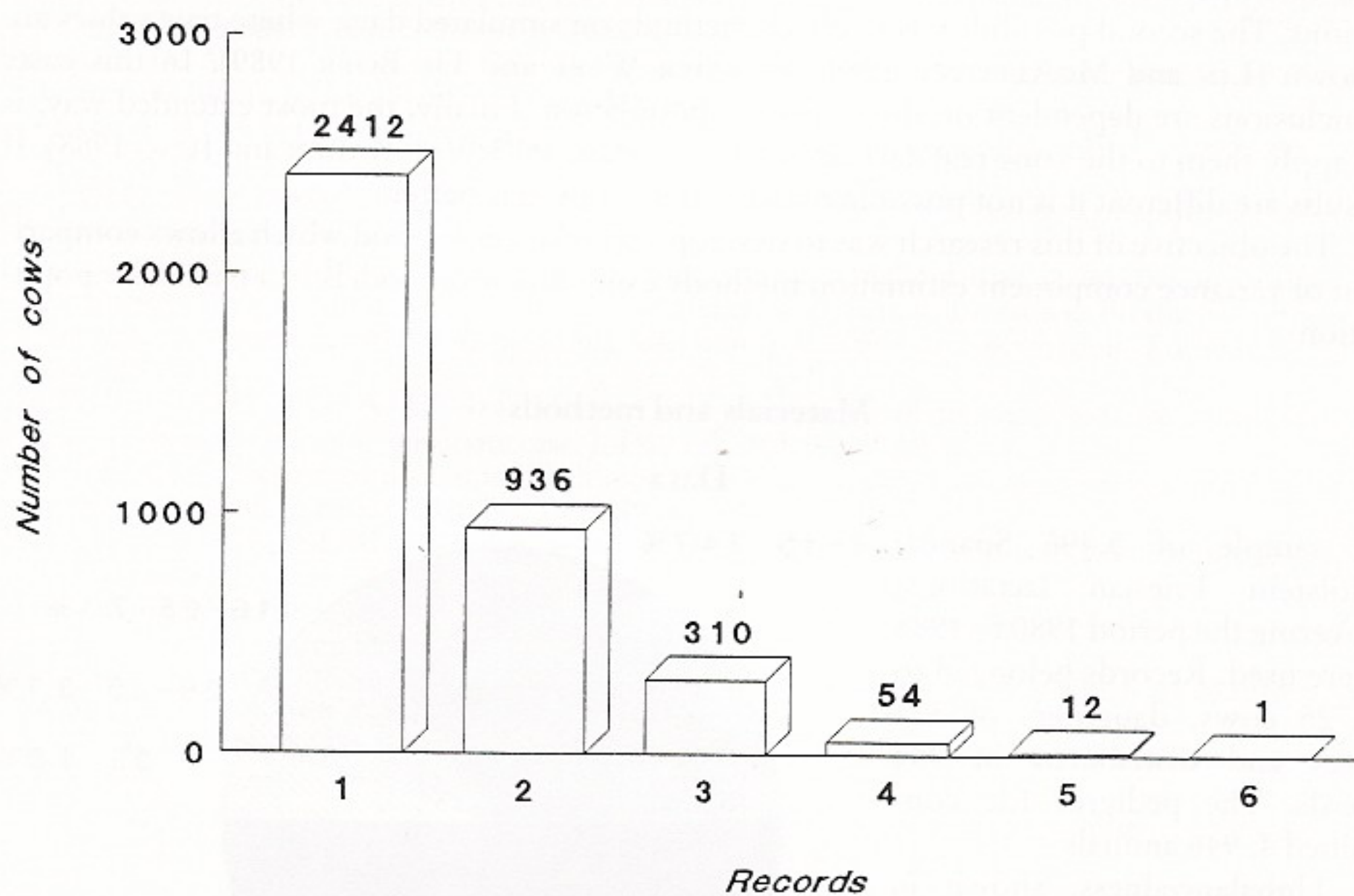


Fig. 3. Distribution of cows according to the number of records

Table 1. Pedigree information

	Number	Percentage
Cows with only the sire known	3157	85 %
Cows with both parents known	568	15 %

- For 85 % of the cows only the sire was known (Table 1).
- 65 % of the cows had only one lactation (Figure 3).

Models

The models tested were:

- Animal model
- Sire and cow nested within sire model

The animal model was:

$$Y_{ijk} = \mu + RA_i + a_j + p_j + e_{ijk}$$

with:

$$E(Y_{ijk}) = \mu + RA_i$$

$$E(a_j) = E(p_j) = E(e_{ijk}) = 0$$

$$\text{Var} \begin{bmatrix} a \\ p \\ e \end{bmatrix} = \begin{bmatrix} A\sigma_a^2 & 0 & 0 \\ 0 & I\sigma_p^2 & 0 \\ 0 & 0 & I\sigma_e^2 \end{bmatrix}$$

where:

- Y_{ijk} is the k^{th} record of the j^{th} cow in the i^{th} herd-year subclass,
 μ the overall mean,
 RA_i the i^{th} fixed herd-year of calving subclass,
 a_j the j^{th} level of the random additive genetic effect,
 p_j the j^{th} level of the random permanent environmental effect of the j^{th} cow
 e_{ijk} the k^{th} level of random residual effect,
 A the numerator relationship matrix,
 σ_a^2 σ_p^2 and σ_e^2 , the genetic, permanent environmental and residual random effects variances, respectively.

The sire model included: overall mean (μ), fixed herd-year effect (RA), random sire effect (s), random cow effect (v) and random residual effect (e):

$$y_{ijkl} = \mu + RA_i + s_j + v_{jk} + e_{ijkl}$$

with expected values and variance-covariance matrix:

$$E(Y_{ijkl}) = \mu + RA_i$$

$$E(s_j) = E(v_{jk}) = E(e_{ijkl}) = 0$$

$$\text{Var} \begin{bmatrix} s \\ v \\ e \end{bmatrix} = \begin{bmatrix} I\sigma_s^2 & 0 & 0 \\ 0 & I\sigma_v^2 & 0 \\ 0 & 0 & I\sigma_e^2 \end{bmatrix}$$

where:

- Y_{ijkl} is the l^{th} record of the k^{th} cow, sired by the j^{th} bull, performing in the i^{th} herd-year
 σ_s^2 , σ_v^2 and σ_e^2 are the sire, cow and residual random effects variances.
 REML procedures were used with both of the models (MEYER 1987a, 1987b, 1988).
 The animal model used sparse matrix solvers (BOLDMAN and VAN VLECK 1991).
 All the runs were carried out on the IBM mainframe model 4381.

Simulation method

Milk and fat yields were simulated three times, using prior estimates of genetic additive, permanent and residual variances, whilst keeping the actual mating scheme, so that the simulation mimicked the actual population.

The simulation method was a modification of Rönningen (1974), allowing fixed effects. Each record in the population was replaced by the addition of the effects in the following animal model:

$$Y_{ijk} = RA_i + a_j + p_j + e_{ijk} \quad [1]$$

where:

RA_i is the i^{th} herd-year fixed effect subclass,

a_j the additive genetic value of the j^{th} cow,

p_j the j^{th} level of the environmental permanent effect,

e_{ijk} the k^{th} residual effect.

The fixed effect value included in the model was the least squared estimation of such effect in the actual sample. All the repetitions were also done without the fixed effect in order to know its influence in the estimation. In this case, model used both for the simulation and for the analysis did not include the herd-year effect.

Variance components were estimated in the actual population using a sire and cow nested model which included the birthyear of the sire as a genetic group effect. Resultant heritabilities and repeatabilities were 0.21 and 0.40 respectively for milk yield and 0.19 and 0.36 for fat yield. These values were used for simulating the random part of the model. Genetic values in the base population were obtained by the product of the genetic standard deviation σ_a , and z_i , a random standard normal value:

$$a_i = z_i \sigma_a$$

Genetic values for animals with known parents were obtained by:

$$a_i = 1/2 a_j + 1/2 a_k + \Phi_i$$

where:

j and k , are the parents of i ,

a_i , a_j and a_k , the genetic values for i , j and k individuals,

Φ_i , the deviation produced by the Mendelian effect in the i^{th} individual.

The permanent environmental p_j and residual e_{ijk} effects were obtained as:

$$p_j = z_j \sigma_p$$

$$e_{ijk} = z_k \sigma_e$$

where z_j and z_k are random standard normal values.

Then we can keep the actual structure of the sample: the same number of daughters per sire, in the same herds, at the same years and with the same number of records per cow.

Results and discussion

Table 2 shows the differences between simulated and obtained values, expressed as a percentage of expected values. Results are displayed for both models, with and without the inclusion of the fixed effect, for the three repetitions, and only for the additive genetic and environmental permanent variances. Results were in general very far from expected values.

Laboratory experiments are designed looking for the optimal structure in order to achieve some purpose (HILL 1970). However, the structure of field data sets can greatly influence the quality of the estimates.

Table 2. Differences between expected and obtained values, expressed as a percentage of the expected values

	MILK				FAT			
	σ_a^2		σ_p^2		σ_a^2		σ_p^2	
	FE*	NFE*	FE	NFE	FE	NFE	FE	NFE
Sire Model	26	-1	13	-18	-30	-31	84	25
	10	2	32	-17	26	-31	38	14
	-11	-2	59	-8	65	19	1	-10
Average	8	0	35	-14	20	-14	41	10
Animal Model	28	1	12	-22	-42	-30	97	24
	18	12	25	-27	19	-29	43	10
	7	2	40	-12	33	17	31	-8
Average	18	5	26	-20	3	-14	57	9

FE = Fixed effects included; NFE = Fixed effects not included.

Phenotypic variance is directly calculated from the data. Residual variance is also obtained in a direct way (HARVILLE 1977). The quality of the estimates for the other two variance components, genetic and permanent variances, depends on the data structure.

Genetic variance estimates could have been biased by the different number of daughters per sire (Figure 2). There are a few sires with many more daughters than the others. However, this disequilibrium influence is corrected when all the animals are connected through the numerator relationship matrix. On the other hand, and looking at Table 1, 85 % of the cows had one parent missing in their pedigree. JENSEN *et al.* (1990) showed that unbiased estimates were achieved only when a complete numerator relationship matrix was included.

Permanent environmental variance estimates were more seriously affected by the unbalanced scheme. 65 % of the cows have only the first record (Figure 3). This estimation would be impossible to carry out if all the animals had produced only once.

Therefore, unbalancedness and missing information could have been the causes of the differences found between obtained and expected values.

The fixed effect, just as it was considered in the simulation, increased phenotypic variance. It will be noted that both genetic and permanent estimates were normally increased when the fixed effect was included (Table 2). Differences between obtained and expected values were also smaller when the fixed effect was ignored.

Although the low number of repetitions does not lead to a definitive conclusion, the results show no advantages of the animal model procedure over the cheaper (in terms of number of equations) sire model for these data, suggesting that the latter may be preferred when data are characterized by a small number of individuals, little pedigree information and highly unbalanced distribution of effects.

Acknowledgements

Authors are extremely grateful to KEITH G. BOLDMAN who modified Meyer program and kindly provided it for this research.

We also appreciate comments by anonymous referee.

Summary

Two models for estimating genetic parameters in animal populations were compared. The cheapest, in terms of computing costs, was based on a sire model and the most expensive on an animal model. In order to check the accuracy of both methods they were applied to the best available sample of Spanish Friesian cattle.

Milk production of each cow, as well as fat contents were simulated using prior estimates of genetic

additive, permanent environment and residual variances, keeping the actual mating scheme, so that the simulation mimicked the actual population structure.

Results, under the above premises, show no advantages of the more costly procedure over the cheaper one, suggesting that a sire model based estimating procedure for genetic parameters may be preferred when a small number of individuals, little pedigree information and highly disequibrated distribution of effects characterize the data.

Zusammenfassung

Vergleich zweier Methoden der Varianzkomponentenschätzung beim Spanischen Holstein

Zwei Modelle zur Schätzung von genetischen Parametern wurden verglichen. Das im Hinblick auf die Rechenleistung günstigere Modell war ein Vatermodell, das aufwendigere Modell ein Tiermodell. Um die Genauigkeit der Modelle in der Praxis zu prüfen, wurden sie auf verfügbares Material der spanischen Holstein-Friesian Population angewendet.

Sowohl die Milchleistung jeder Kuh als auch der Fettgehalt wurden simuliert mit Hilfe von Schätzwerten der additiv-genetischen Varianz, der permanenten und der residuellen Umweltvarianz. Dabei wurde das tatsächliche Ungleichgewicht in der Population berücksichtigt.

Unter den obengenannten Voraussetzungen zeigen die Ergebnisse keinen Vorteil der rechnerisch aufwendigeren Prozedur gegenüber der einfacheren. Die Schlußfolgerung lautet, daß das Vatermodell unter den Bedingungen der kleinen Tierzahlen, wenig Pedigree-Informationen und einer stark unbalancierten Verteilung der Tiere auf die einzelnen Effektklassen zu bevorzugen ist.

References

- DEMPFLE, L.; HAGGER, C.; SCHNEEBERGER, M., 1983: On the estimation of genetic parameters via variance components. *Génét. Sél. Evol.* **15**, 425-444.
- BOLDMAN, K. G.; VAN VLECK, L. D., 1991: Derivate-free Restricted Maximum Likelihood in Animal Models with a Sparse Matrix Solver. *J. Dairy Sci.*, **74**, 4337-4343.
- HARVILLE, D. A., 1977: Maximum likelihood approaches to variance component estimation and to related problems. *J. An., Statist. Assoc.* **72**, 320-338.
- HENDERSON, C. R., 1975: Comparison of alternative sire evaluation methods. *J. Anim. Sci.* **45**, 760-770.
- HILL, W. G., 1970: Design of experiments to estimate heritability by regression of offspring on selected parents. *Biometrics* **26**, 566-571.
- JENSEN, J.; MOORE, T. L.; MAO, I. L., 1990: Sampling of traits and data in estimation of genetic parameters. In: *Proc. 4th, World Congress, on Genetics applied, to Livestock, Production. Plenary lectures, molecular genetics and mapping, selection, prediction and estimation*. Edinburgh, UK. 281-284.
- LIN, C. Y.; MCALLISTER, A. J., 1984: Monte Carlo comparison of four methods for estimation of genetic parameters in the univariate case. *J. Dairy Sci.* **67**, 2389-2398.
- MEYER, K., 1987: Restricted maximum likelihood to estimate variance components for mixed models with two random factors. *Génét. Sél. Evol.* **19**, 49-68.
- MEYER, K., 1988a: DFRML-a set of programs to estimate variance components under an individual animal model. *J. Dairy Sci.* **71**, (suppl. 2), 33-34.
- MEYER, K., 1988b: DFRML. Programs to estimate variance components for individual animal models by restricted maximum likelihood. User notes. Univ. of Edinburgh.
- RÖNNINGEN, K., 1974: Monte Carlo Simulation of Statistical-Biological Models which are of Interest in Animal Breeding. *Act. Agric. Scan.* **24**, 135-142.
- VAN DER WERF, J. H. J.; DE BOER, W., 1989: Influence of nonadditive effects on estimations of genetic parameters in dairy cattle. *J. Dairy Sci.* **72**, 2606-2614.
- VAN RADEN, P. M.; JUNG, Y. C., 1988: A general purpose approximation to restricted maximum likelihood: The tilde-hat approach. *J. Dairy Sci.* **71**, 187-194.

Authors' address: Dr. J. P. GUTIÉRREZ, Faculty of Veterinary Medicine, University Complutense, E-28040 Madrid, Spain